Package ‘GOTHiC’

April 25, 2024

Title Binomial test for Hi-C data analysis

Description This is a Hi-C analysis package using a cumulative binomial test to detect interactions between distal genomic loci that have significantly more reads than expected by chance in Hi-C experiments. It takes mapped paired NGS reads as input and gives back the list of significant interactions for a given bin size in the genome.

Version 1.39.0

Date 2021-11-18

Author Borbala Mifsud and Robert Sugar

Maintainer Borbala Mifsud <b.mifsud@qmul.ac.uk>

Depends R (>= 3.5.0), methods, GenomicRanges, Biostrings, BSgenome, data.table

Imports BiocGenerics, S4Vectors (>= 0.9.38), IRanges, Rsamtools, ShortRead, rtracklayer, ggplot2, BiocManager, grDevices, utils, stats, GenomeInfoDb

Suggests HiCDataLymphoblast

Enhances parallel

License GPL-3

biocViews ImmunoOncology, Sequencing, Preprocessing, Epigenetics, HiC

git_url https://git.bioconductor.org/packages/GOTHiC

git_branch devel

git_last_commit f393257

git_last_commit_date 2023-10-24

Repository Bioconductor 3.19

Date/Publication 2024-04-25

Contents

filtered 2
GOTHiC 3
GOTHiChicup 5
mapReadsToRestrictionSites 6
pairReads 8

Index **10**

`filtered` *A GenomicRangesList object used as an example in the GOTHiC package*

Description

`filtered` is a GenomicRangesList example object used as an example for the binomialHiC package. This GenomicRangesList contains reads from a human lymphoblastoid cell line HiC experiment (Lieberman-Aiden et al. 2009) for chr20, that were mapped to the genome, paired and PCR duplicate-filtered.

Usage

```
data(lymphoid_chr20_paired_filtered)
```

Format

The format is: GenomicRangesList with 2 slots: $paired_reads_1 contains the coordinates for one end of the paired reads $paired_reads_2 contains the coordinates for the other end of the paired reads

Author(s)

Borbala Gerle and Robert Sugar

See Also

`mapReadsToRestrictionSites`

Examples

```
data(lymphoid_chr20_paired_filtered)
```

GOTHiC *Genome Organisation Through HiC*

Description

GOTHiC performs a cumulative binomial test to detect interactions between distal genomic loci that have significantly more reads than expected by chance in Hi-C experiments. It takes mapped paired NGS reads as input and gives back the list of significant interactions for a given bin size in the genome.

Usage

```
GOTHiC(fileName1, fileName2, sampleName, res,
BSgenomeName='BSgenome.Hsapiens.UCSC.hg19',
genome=BSgenome.Hsapiens.UCSC.hg19, restrictionSite='A^AGCTT',
enzyme='HindIII',cistrans='all',filterdist=10000,
DUPLICATETHRESHOLD=1, fileType='BAM', parallel=FALSE, cores=NULL)
```

Arguments

fileName1	File containing the mapped reads of the first fragment ends (BAM or Bowtie format)
fileName2	File containing the mapped reads of the second fragment ends (BAM or Bowtie format)
sampleName	A character string that will be used to name the exported BedGraph file containing the coverage, R object files with paired and mapped reads, and the final data frame with the results from the binomial test. They will be saved in the current directory.
res	An integer that gives the required bin size or resolution of the contact map e.g. 1000000.
BSgenomeName	A character string of the BSgenome package required to make the restriction fragment file containing information for both the organism the experiment was made in, and the genome version the reads were mapped to. The default is the current human genome build 'BSgenome.Hsapiens.UCSC.hg19'.
genome	The BSgenome package required to make the restriction fragment file containing information for both the organism the experiment was made in, and the genome version the reads were mapped to. The default is the current human genome build BSgenome.Hsapiens.UCSC.hg19.
restrictionSite	A character string that specifies the enzymes recognition site, ^ indicating where the enzyme actually cuts. The default is the HindIII restriction site: 'A^AGCTT'.
enzyme	A character string containing the name of the enzyme used during the Hi-C experiment (i.e. "HindIII", "NcoI"). The default is "HindIII".

`cistrans` A character string with three possibilities. "all" runs the binomial test on all interactions, "cis" runs the binomial test only on intrachromosomal/cis interactions, "trans" runs the binomial test only on interchromosomal/trans interactions.

`filterdist` An integer specifying the distance between the midpoint of fragments under which interactions are filtered out in order to filter for those read-pairs where the digestion was incomplete. The default is 10000.

`DUPLICATETHRESHOLD` An integer specifying the maximum amount of duplicated paired-end reads allowed, over that value it is expected to be PCR bias. The default is 1.

`fileType` A character string specifying the format of the aligned reads. The default is 'BAM'. Other accepted format is 'Bowtie'.

`parallel` Logical argument. If TRUE the mapping and the binomial test will be performed faster using multiple cores. The default is FALSE.

`cores` An integer specifying the number of cores used in the parallel processing if parellel=TRUE. The default is NULL.

Value

A data.frame containing elements

`chr1 / chr2` chromosome(s) containing interacting regions 1 and 2

`locus1 / locus2` start positions of the interacting regions 1 and 2 in the corresponding chromosome(s)

`relCoverage1 / relCoverage2` relative coverage corresponding to regions 1 and 2

`probability` expected frequency

`expected` expected number of reads

`readCount` observed reads number

`pvalue` binomial p-value

`qvalue` binomial p-value corrected for multi-testing with Benjamini-Hochberg

`logObservedOverExpected` observed/expected read numbers log ratio

Author(s)

Borbala Mifsud and Robert Sugar

See Also

`binom.test`, `pairReads`, `mapReadsToRestrictionSites`

Examples

```
library(GOTHiC)
dirPath <- system.file("extdata", package="HiCDataLymphoblast")
fileName1 <- list.files(dirPath, full.names=TRUE)[1]
fileName2 <- list.files(dirPath, full.names=TRUE)[2]
binom=GOTHiC(fileName1, fileName2, sampleName='lymphoid_chr20', res=1000000,
BSgenomeName='BSgenome.Hsapiens.UCSC.hg18', genome=BSgenome.Hsapiens.UCSC.hg18,
restrictionSite='A^AGCTT', enzyme='HindIII',cistrans='all', filterdist=10000,
DUPLICATETHRESHOLD=1, fileType='Table', parallel=FALSE, cores=NULL)
```

GOTHiChicup *Genome Organisation Through HiC from HiCUP output*

Description

GOTHiChicup performs a cumulative binomial test to detect interactions between distal genomic loci that have significantly more reads than expected by chance in Hi-C experiments. It takes mapped and filtered paired NGS reads from HiCUP as input and gives back the list of significant interactions for a given bin size in the genome.

Usage

```
GOTHiChicup(fileName, sampleName, res, restrictionFile, cistrans='all', parallel=FALSE, cores=NULL)
```

Arguments

fileName	A character string with the name of the file containing the mapped, filtered reads from HiCUP, after the default HiCUP output is converted to a table containing only the first 4 columns (read ID, flag, chromosome and start positions). Can be gzipped. (Tab separated text format)
sampleName	A character string that will be used to name the quality control plot. It will be saved in the current directory.
res	An integer that gives the required bin size or resolution of the contact map e.g. 1000000, for fragment level use 1.
restrictionFile	A character string with the name of the digest file from HiCUP. It is used to map reads to restriction fragments. (.txt file name)
cistrans	A character string with three possibilities. "all" runs the binomial test on all interactions, "cis" runs the binomial test only on intrachromosomal/cis interactions, "trans" runs the binomial test only on interchromosomal/trans interactions.
parallel	Logical argument. If TRUE the mapping and the binomial test will be performed faster using multiple cores. The default is FALSE.
cores	An integer specifying the number of cores used in the parallel processing if parellel=TRUE. The default is NULL.

Value

A data.frame containing elements

- `chr1 / chr2` chromosome(s) containing interacting regions 1 and 2
- `locus1 / locus2` start positions of the interacting regions 1 and 2 in the corresponding chromosome(s)
- `relCoverage1 / relCoverage2` relative coverage corresponding to regions 1 and 2
- `probability` expected frequency
- `expected` expected number of reads
- `readCount` observed reads number
- `pvalue` binomial p-value
- `qvalue` binomial p-value corrected for multi-testing with Benjamini-Hochberg
- `logObservedOverExpected` observed/expected read numbers log ratio

Author(s)

Borbala Mifsud and Robert Sugar

See Also

`binom.test`

Examples

```
library(GOTHiC)
dirPath <- system.file("extdata", package="HiCDataLymphoblast")
fileName <- list.files(dirPath, full.names=TRUE)[4]
restrictionFile <- list.files(dirPath, full.names=TRUE)[3]
binom=GOTHiChicup(fileName, sampleName='lymphoid_chr20', res=1000000,
restrictionFile, cistrans='all', parallel=FALSE, cores=NULL)
```

---

mapReadsToRestrictionSites

Function to map aligned and paired reads to the restriction fragments

---

Description

This function takes mapped paired NGS reads in the format of a GenomicRangesList object where the two end of the reads are in the GenomicRanges paired_reads_1 and paired_reads_2. It prepares the digestion file from the genome supplied to it with the given restriction enzyme and specificity and maps the reads to the fragments.

Usage

```
mapReadsToRestrictionSites(pairedReadsFile, sampleName,
BSgenomeName, genome, restrictionSite, enzyme, parallel=F, cores=1)
```

Arguments

`pairedReadsFile`
: R object of GenomicRangesList containing paired_reads_1 and paired_reads_2 GenomicRanges with the paired mapped reads from a Hi-C experiment.

`sampleName`
: A character string that will be used to name the exported R object file with the mapped reads containing a GenomicRangesList with slots locus1 and locus2. It will be saved in the current directory.

`BSgenomeName`
: A character string of the BSgenome package required to make the restriction fragment file containing information for both the organism the experiment was made in, and the genome version the reads were mapped to. The default is the current human genome build 'BSgenome.Hsapiens.UCSC.hg19'.

`genome`
: The BSgenome package required to make the restriction fragment file containing information for both the organism the experiment was made in, and the genome version the reads were mapped to. The default is the current human genome build BSgenome.Hsapiens.UCSC.hg19.

`restrictionSite`
: A character string that specifies the enzymes recognition site, ^ indicating where the enzyme actually cuts. The default is the HindIII restriction site: 'A^AGCTT'.

`enzyme`
: A character string containing the name of the enzyme used during the Hi-C experiment (i.e. "HindIII", "NcoI"). The default is "HindIII".

`parallel`
: Logical argument. If TRUE the mapping will be performed faster using multiple cores. The default is FALSE.

`cores`
: An integer specifying the number of cores used in the parallel processing if parellel=TRUE. The default is 1.

Value

A GenomicRangesList

`locus1`
: GenomicRanges with the coordinates of the start of the fragment where one end of the read mapped

`locus2`
: GenomicRanges with the coordinates of the start of the fragment where the other end of the read mapped

Author(s)

Borbala Mifsud and Robert Sugar

See Also

`pairReads`, `GOTHiC`

Examples

```
library(GOTHiC)
data(lymphoid_chr20_paired_filtered)
mapped=mapReadsToRestrictionSites(filtered, sampleName='lymphoid_chr20',
BSgenomeName='BSgenome.Hsapiens.UCSC.hg18', genome=BSgenome.Hsapiens.UCSC.hg18,
restrictionSite='A^AGCTT', enzyme='HindIII', parallel=FALSE, cores=1)
```

---

pairReads *Function pairs aligned paired NGS reads*

Description

This function takes bowtie output files, pairs the reads, only keeps those where both ends mapped, filters for perfect duplicates to avoid PCR bias, and saves and returns a GenomicRangesList object that contains the paired_reads_1 and paired_reads_2 GenomicRanges with the paired reads

Usage

```
pairReads(fileName1, fileName2, sampleName, DUPLICATETHRESHOLD = 1,
fileType='BAM')
```

Arguments

`fileName1`	File containing the mapped reads of the first fragment ends (BAM or Bowtie format)
`fileName2`	File containing the mapped reads of the second fragment ends (BAM or Bowtie format)
`sampleName`	A character string that will be used to name the exported BedGraph file containing the coverage, and the R object file with paired reads. They will be saved in the current directory.
`DUPLICATETHRESHOLD`	An integer specifying the maximum amount of duplicated paired-end reads allowed, over that value it is expected to be PCR bias. The default is 1.
`fileType`	A character string specifying the format of the aligned reads. The default is 'BAM'. Other accepted format is 'Bowtie'.

Value

A GenomicRangesList called filtered

`paired_reads_1`	GenomicRanges with the coordinates of where one end of the read mapped
`paired_reads_2`	GenomicRanges with the coordinates of where the other end of the read mapped

Author(s)

Borbala Mifsud and Robert Sugar

See Also

mapReadsToRestrictionSites, GOTHiC

Examples

```
library(GOTHiC)
dirPath <- system.file("extdata", package="HiCDataLymphoblast")
fileName1 <- list.files(dirPath, full.names=TRUE)[1]
fileName2 <- list.files(dirPath, full.names=TRUE)[2]
paired <- pairReads(fileName1, fileName2, sampleName='lymphoid_chr20',
DUPLICATETHRESHOLD = 1, fileType='Table')
```

Index

∗ **datasets**
    filtered, 2
∗ **manip**
    GOTHiC, 3
    GOTHiChicup, 5
    mapReadsToRestrictionSites, 6
    pairReads, 8
∗ **package**
    GOTHiC, 3
    GOTHiChicup, 5

filtered, 2

GOTHiC, 3
GOTHiChicup, 5

lymphoid_chr20_paired_filtered
    (filtered), 2

mapReadsToRestrictionSites, 6

pairReads, 8